ARTICLE IN PRESS IJA:7619

International Journal of Approximate Reasoning ●●● (●●●●) ●●●–●●●

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Refining a Bayesian Network using a Chain Event Graph ☆

L.M. Barclay *, J.L. Hutton, J.Q. Smith

Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom

ARTICLE INFO

Article history:
Received 13 December 2012
Received in revised form 3 April 2013
Accepted 14 May 2013
Available online xxxx

Keywords:
Chain Event Graphs
Bayesian Networks
Bayesian model selection
Causality

ABSTRACT

The search for a useful explanatory model based on a Bayesian Network (BN) now has a long and successful history. However, when the dependence structure between the variables of the problem is asymmetric then this cannot be captured by the BN. The Chain Event Graph (CEG) provides a richer class of models which incorporates these types of dependence structures as well as retaining the property that conclusions can be easily read back to the client. We demonstrate on a real health study how the CEG leads us to promising higher scoring models and further enables us to make more refined conclusions than can be made from the BN. Further we show how these graphs can express causal hypotheses about possible interventions that could be enforced.

1. Introduction

The Bayesian Network (BN) is the most widely used graphical model [7,16] which expresses the relationship between the variables of the system in terms of conditional independence statements. Its graphical structure makes it a particularly useful tool to feed conclusions back to the client and it has therefore been employed in many real-world applications. However, in certain cases the BN does not provide a rich enough structure to incorporate all information obtainable from the data set. This is the case, for example, when the conditional independence statements of the problem are asymmetric or only certain combinations of variables affect another variable and this cannot be represented simply by the directed edges between variables in the BN [21]. To take these features into account extensions to the BN have been proposed, mostly in the form of tables or tree-like structures which are added to the graph, leading to the context-specific Bayesian Network [5,21,11]. However, these methods focus primarily on efficient propagation and learning and lose the benefit of the BN's expressiveness for the client. An interesting related graphical model is the Recursive Probability Tree [6] which also focuses on efficient computation of context-specific independencies.

The Chain Event Graph (CEG) is a new flexible class of graphical models which can represent asymmetric structures directly in its topology. It is related to the Probability Decision Graph (PDG) [19,15,18] and retains the framework of a probability tree in a more compact graph. Because of its graphical derivation it inherits many of the benefits of a BN. For example, we can read off conditional independence statements directly from the topology of the graph [24,26], carry out model selection on CEGs [10] and run fast propagation algorithms [27]. It also admits a causal extension [28,25]. However, it is a more general class than either the PDG or the discrete BN. While the PDG and BN are proven to be distinct classes of models, neither containing the other [14], the CEG contains each of these classes as a special case [24]. Therefore, it is very straightforward to exploit and develop the technologies originally designed for the BN to this much richer class of CEGs. In this paper we illustrate how we use the CEG to refine an initial BN model description, demonstrating how this can provide a

* Corresponding author.
E-mail addresses: L.M.Barclay@warwick.ac.uk (L.M. Barclay), J.L.Hutton@warwick.ac.uk (J.L. Hutton), J.Q.Smith@warwick.ac.uk (J.Q. Smith).

0888-613X/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2013.05.006

Please cite this article in press as: L.M. Barclay et al., Refining a Bayesian Network using a Chain Event Graph, International Journal of Approximate Reasoning (2013), http://dx.doi.org/10.1016/j.ijar.2013.05.006

more detailed but still transparent explanation of an underlying process, which seamlessly enhances an original BN analysis. The application in this paper explores a particular study of the Christchurch Health and Development Study [1], looking at the effect of social and family factors on children's health in a New Zealand birth cohort [9].

We begin the paper with a description of the programme of study forming the basis of this analysis. In Section 3 we report how an initial routine BN model search enabled us to elaborate the conclusions made in the original study. In Sections 4 and 5 we show how we then made more detailed exploration using the CEG, discovering a much better fitting model that – like a BN – could be read back to the client in narrative style. We also illustrate in Section 6 how the fitted model can be given a causal interpretation which, if valid, allows us to perform various 'what–if' analyses under various policy controls. In the final section we propose further extensions to the present analysis.

2. The Christchurch Health and Development Study

The Christchurch Health and Development Study (CHDS) is carried out by a research group at Otago University, led by Professor David Fergusson. It is a cohort study which has followed up 1265 children born in mid 1977 in Christchurch, New Zealand, for over 30 years. The CHDS has explored the children's development from childhood to adulthood regarding their education, behaviour and health with respect to a wide range of social, economic and family factors.

The running example used in this paper reanalyses an early subset of the CHDS discussed in [9] which studies the first five years of the Christchurch cohort looking at the effect the family's social background, the economic status and the number of family life events have on the child's health which is measured by rates of hospital admission.

Based on previous study of the data set, Fergusson et al. [9] concluded to only consider admissions due to illness and accidents as these were the only reasons for admissions that are sensitive to social and family situation. To describe the family's social background the CHDS group collected information about the mother's education and age at birth, the family's socioeconomic status and ethnic origin, and whether the child grew up in a single or two parent family. These variables were then combined using factor analysis to give a single measure of the social background (see [8] for details). Similarly, the economic status was measured as a function of the family income, possible financial difficulties and by rating the standard of living and the quality of the accommodation of the child. Again these were simplified into a single measure of the overall economic situation. Of particular interest in this study was whether the effect of adverse life events in the children's lives might be associated with an increased rate of illness. This has provoked controversy and continues to be a subject of research [29]. Twenty events were classed as life events, based on a variation of the Holmes and Rahe Social Readjustment Rating Scale [13], in which the mother of the child was interviewed and reported on the events that occurred. These included the experience of moving house, the husband changing job, the death of a close friend or relative, serious financial problems within the family, divorce, or a serious illness or accident within the family [3].

Data was collected through a variety of sources: Mothers were interviewed every year on the family life events, while the economic situation was predominantly rated every year by an interviewer. Hospital admissions were also obtained by interviewing the mother of every child at each year and further by asking them to keep a health diary record. The results were additionally compared with the central records of the Christchurch Hospital.

Rates of hospital admissions were first compared for the various levels of each covariate separately using one-way analysis of variance, concluding that the more socially or economically disadvantaged the child's background (without adjusting for the other factors) the higher the hospital admission rate tended to be. In addition to this the probability of an admission increased significantly with the number of family life events. For further analysis a Cox proportional hazards model estimating the risk of hospital admission over the five years was fitted. Results from [9] showed that according to this model the family's economic status did not influence the risk of admission significantly once adjusting for the other covariates, suggesting that in this type of population financial problems were not the main reason for health problems. In contrast, family life events and social background both appeared to have a significant impact on the admission rates, even once having adjusted for the other covariates. In particular, the most significant association was found between the hospital admissions rate and the number of family life events. These interesting conclusions motivated the following analysis. We demonstrate that our graphical methods not only allow us to disentangle the association between the variables but enable us to present our results in a transparent way.

Complete data was available to us for 890 children and so our analysis was carried out on these. To construct the four variables of interest we aimed to follow as far as possible the methodology of Fergusson et al. [9]. However, as the variables describing the social and economic background are discrete, predominantly with few categories, we slightly adapted the methods of Fergusson et al. [9], who use a factor analysis. Instead we fitted a latent-class model to construct a categorical variable for the social background and the economic situation. The predicted classes of each child were used to describe its social and economic situation. For simplicity, we here assumed binary latent classes throughout, distinguishing between a high or low social background and a high or low economic situation. We further made a distinction between 'no hospital admission' and 'at least one hospital admission' since the counts for more than one admission are sparse. Similarly, we divided the life events into the three approximately equal sized categories: 'low' (0–5 events over the five years), 'average' (6–9 events) and 'high' (at least 10 events). There were 169 (19%) children with at least one admission and the proportion of admission overall ranged from 0.12 to 0.26 (Table 1).

Please cite this article in press as: L.M. Barclay et al., Refining a Bayesian Network using a Chain Event Graph, International Journal of Approximate Reasoning (2013), http://dx.doi.org/10.1016/j.ijar.2013.05.006

Table 1
Summary statistics and proportions of admission.

Admissions	No admission 721	$\geqslant 1$ admission 169	
Social background	High 507 (0.148)	Low 383 (0.245)	
Economic situation	High 283 (0.148)	Low 607 (0.209)	
Number of life events	0–5 events 329 (0.119)	6–9 events 295 (0.210)	$\geqslant 10$ events 266 (0.256)

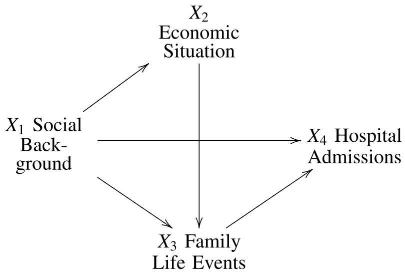

Fig. 1. DAG for the effects on hospital admission.

3. Translation into a Bayesian Network versus model selection

We began by fitting a standard BN to this application outlined in the previous section.
We let

- $X_1 =$ family social background
- $X_2 =$ family economic situation
- $X_3 =$ number of family life events
- $X_4 =$ hospital admission.

The main conclusions of Fergusson et al. [9] could then be expressed as the single conditional independence statement $X_4 \perp\!\!\!\perp X_2 \mid X_1, X_3$ or equivalently by the DAG given in Fig. 1. Note that the conditional independence statement could be represented by several DAGs (for example the edges $X_1 \to X_4$, $X_3 \to X_4$ and $X_2 \to X_3$ could be reversed). However, X_1 is measured at birth and further we are predominantly interested in the effect of the other variables on the hospital admission. Therefore, we chose the DAG that also had a plausible causal interpretation. We discuss the causal hypotheses that can be made from this structure further at the end of the section.

Our first step was to carry out a standard model selection procedure to find the Maximum a Posteriori (MAP) BN structure over the four variables. The network structures were scored using the Bayesian Dirichlet (BD) metric in [12]. As recommended by [17] the prior Dirichlet distributions were given an equivalent sample size of 3, the highest number of categories taken by a variable in the problem. For transparency, we chose a prior network such that the distribution over all possible configurations was uniform and assumed that structures were a priori equally likely. We use Bayes Factor (BF) scores relative to the MAP BN model throughout for the comparison of different models although other scoring methods could have been used.

An exhaustive search using the 'deal' package in R [4] over all possible BNs on the four variables found the MAP model to be the DAG given in Fig. 2(a) with the corresponding set of conditional independence statements given by: $X_3 \perp\!\!\!\perp X_2 \mid X_1$, $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$ and the associated predictive Conditional Probability Table (CPT) given in Table 2.

In comparison with the network structure derived from [9] the MAP model found was more sparse: It suggests an additional conditional independency between the economic situation and the family life events given the social background ($X_3 \perp\!\!\!\perp X_2 \mid X_1$) and further expresses a direct dependency only between the life events and the hospital admissions and not between social background and admissions ($X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$). Further, the exhaustive search over all possible structures revealed three further network structures scoring only slightly less than the MAP model, given in Fig. 2 (b)–(d). Network structure (b) swaps the directed edge from family life events to admissions with an edge from the social background to the admissions. Structure (c) introduces an extra edge between the economic situation and the family life events and model (d) again exchanges the edge of model (c) from family life events to admissions for an edge between social background and admissions. Although the closeness of scores of different competing BNs might be due to sparseness of the data set, it is also mildly suggestive that a model that combines features of different competing BNs is the actual generating process. We demonstrate below that this second possibility is strongly indicated.

Please cite this article in press as: L.M. Barclay et al., Refining a Bayesian Network using a Chain Event Graph, International Journal of Approximate Reasoning (2013), http://dx.doi.org/10.1016/j.ijar.2013.05.006

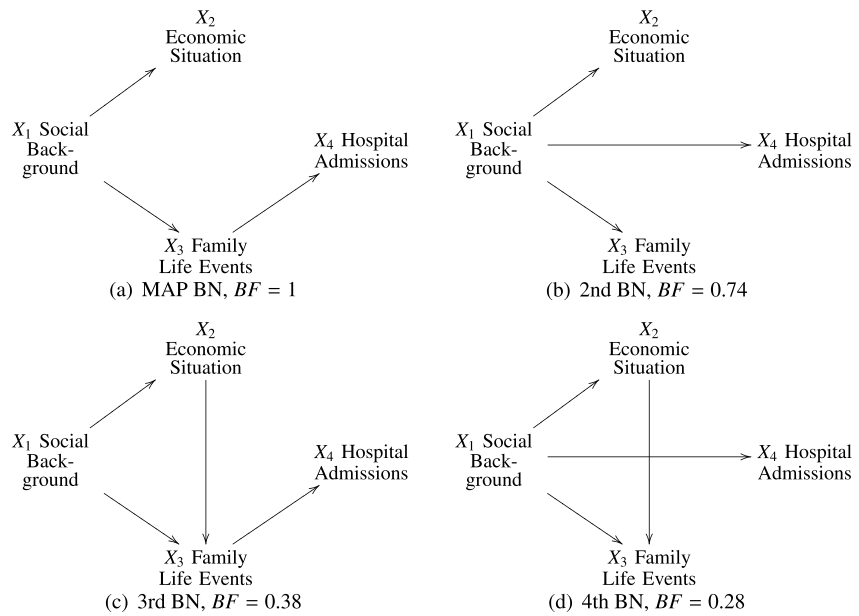

Fig. 2. Highest scoring BN structures.

Table 2
Conditional probability table for MAP BN.

Conditional probability vector	
$(P(X_1 = \text{High}), P(X_1 = \text{Low}))$	$(0.569, 0.431)$
$(P(X_2 = \text{High}\mid X_1 = \text{High}), P(X_2 = \text{Low}\mid X_1 = \text{High}))$	$(0.468, 0.532)$
$(P(X_2 = \text{High}\mid X_1 = \text{Low}) P(X_2 = \text{Low}\mid X_1 = \text{Low}))$	$(0.122, 0.878)$
$(P(X_3 = \text{Low}\mid X_1 = \text{High}), P(X_3 = \text{Average}\mid X_1 = \text{High}), P(X_3 = \text{High}\mid X_1 = \text{High}))$	$(0.461, 0.347, 0.192)$
$(P(X_3 = \text{Low}\mid X_1 = \text{Low}), P(X_3 = \text{Average}\mid X_1 = \text{Low}), P(X_3 = \text{High}\mid X_1 = \text{Low}))$	$(0.248, 0.311, 0.441)$
$(P(X_4 = \text{No admission}\mid X_3 = \text{Low}), P(X_4 = \text{Admission}\mid X_3 = \text{Low}))$	$(0.881, 0.119)$
$(P(X_4 = \text{No admission}\mid X_3 = \text{Average}), P(X_4 = \text{Admission}\mid X_3 = \text{Average}))$	$(0.789, 0.211)$
$(P(X_4 = \text{No admission}\mid X_3 = \text{High}), P(X_4 = \text{Admission}\mid X_3 = \text{High}))$	$(0.744, 0.256)$

In these applications we are often interested in the causal effect of social background, economic situation and family life events on hospital admissions. For example, if it were possible to intervene on the number of life events and enact a policy ensuring that the life events on a particular unit would always lie in the 'low' category, then we could conclude from the BN and its associated CPT that the consequent probability of hospital admission would be reduced to 0.119 (compare Table 2). However, these types of interventions may also be asymmetric. We may be interested in the effect of an intervention which gives only families from a low social background financial aid, or similarly, we would like to know the predicted probability of an admission if we could stop children from a low social background having a high number of life events. These types of interventions can be simply represented within a CEG, which we describe in the next two sections.

4. Refining a Bayesian Network using a Chain Event Graph

We therefore introduce a new, more flexible, class of models called the Chain Event Graphs (CEG) [24,28]. It has several advantages over the BN: Firstly, it is an enhancement of the standard BN by allowing for asymmetries within the dependence structure as well as capturing discrete BNs as a particular subclass. Secondly, the CEG is derived from a probability tree retaining its paths in a more compact graph. It hence enables us to provide a plausible story of the way in which different factors affect children's health, sharing therefore with the BN the property of providing an evocative graphical framework through which conclusions can be read back to the client, who in this setting would be an advisor to policy makers within the social services.

The CEG is derived from a probability tree which is simplified into a CEG by introducing the concepts of 'stages' and 'positions'. These group the vertices in the tree together according to the associated conditional probabilities on their edges. We say that two vertices are in the same stage u, when we have a one-to-one mapping between the edges emanating from the two vertices and their associated conditional probabilities are the same. The full set of stages is denoted by $J(T)$. When two vertices are in the same stage then their associated edges are coloured such that corresponding edges have the same colour. The resulting tree is called a staged tree. For clarity, we here use instead of colours different node shapes when two

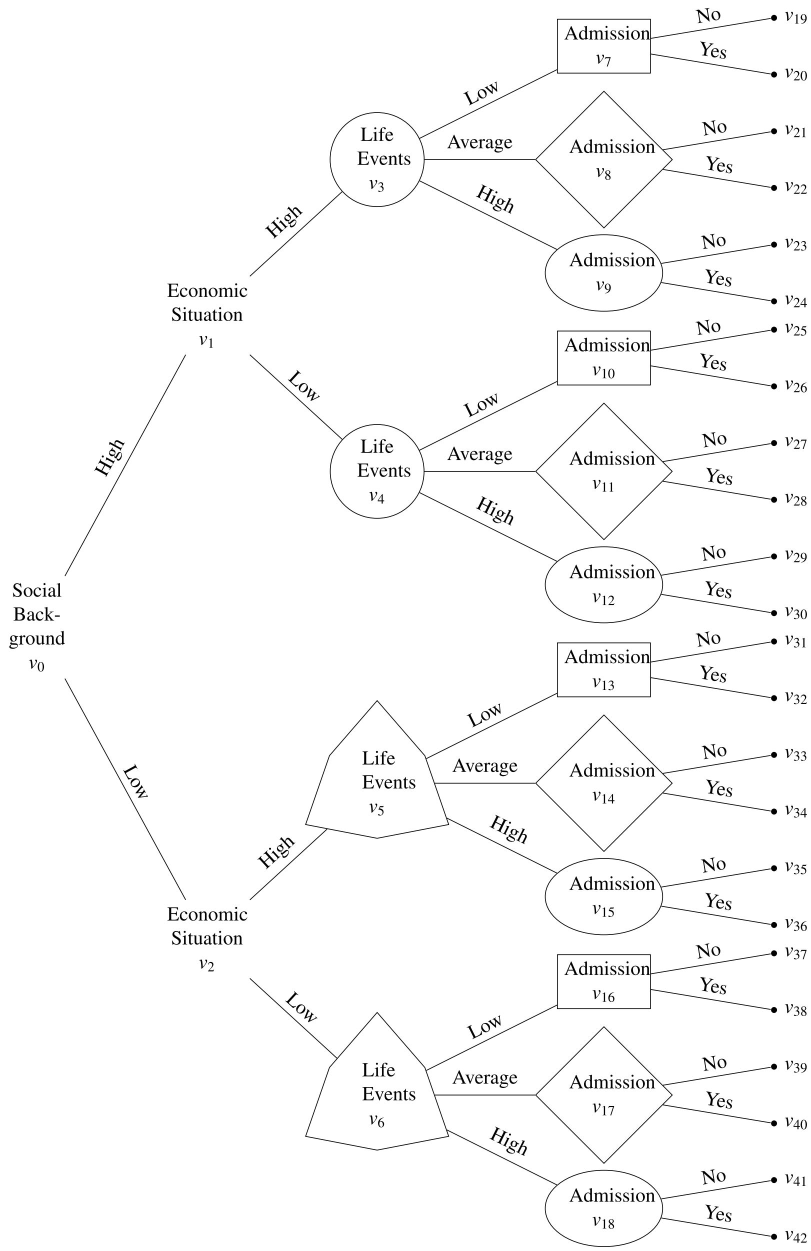

Fig. 3. Staged tree.

vertices are in the same stage. A finer partition of the vertices is given by the concept of positions. We say that two vertices are in the same position w when their full subtrees and all associated probabilities are the same.

The CEG is then constructed from the staged tree where each position is represented by a single vertex, and the set of leaf-nodes are collected in a single position, called w_∞. Hence it is a graphical model whose vertices equal the positions of the corresponding staged tree. Its edge set is defined as follows: For each position w a single representative vertex $v(w)$ is chosen. Then there exists an edge from a position w to another position w' in the CEG for every edge in the staged tree from $v(w)$ to a vertex $v' \in w'$. As a stage may be made up of several positions, the stages of the tree are indicated by adding an undirected dotted line between any two positions that are in the same stage.

We illustrate below how a discrete BN can be represented as a CEG using the MAP BN structure given in Fig. 2(a). We can first draw a tree corresponding to our discrete BN such that parent variables appear before their children in the ordering of the tree (Fig. 3). Note that this ordering is not necessarily unique as a set of conditional independence statements can be represented by several BN structures. Further given a structure the ordering may only be partial, allowing certain parent variables to be interchanged. Here, we decide on the ordering $\boldsymbol{X} = (X_1, X_2, X_3, X_4)$. Putting the social background as our

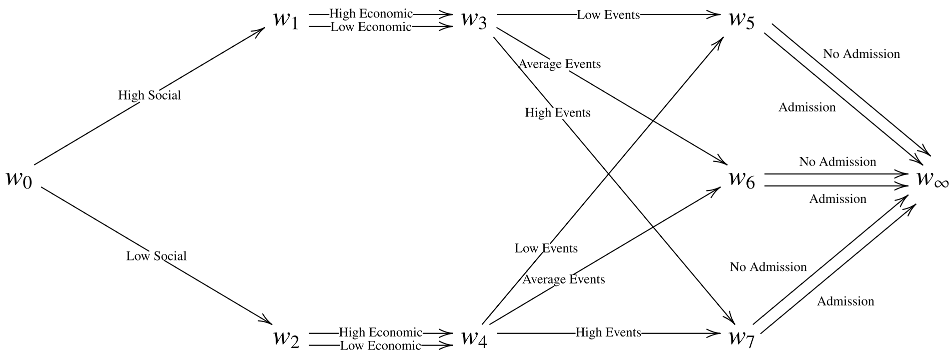

Fig. 4. CEG derived from MAP BN structure.

first variable is an obvious choice, as it is measured only at birth and hence cannot be affected by variables measured after birth. The admissions are put as the final variable as we are interested in the effect that the other three variables have on it. The economic situation is placed before the life events, which suggests that the economic situation may affect the number of life events (e.g. change of job, financial problems). However, it is also plausible that the life events affect the economic situation. We will discuss later the effect of switching the economic situation and the life events in the probability tree.

The conditional independence statements of a faithful BN can then be uniquely represented by defining stages on the tree. For a detailed proof that every discrete BN can be written as a CEG see [24] (Section 3.2, p. 56). Note that a conditional independency in the BN is of the form

$$X_i \perp\!\!\!\perp \{X_1, \ldots, X_{i-1}\} \backslash pa(X_i) \mid pa(X_i),$$

where $pa(X_i)$ is the parent set of X_i. To represent this in terms of stages in the tree we put all the vertices describing X_i into the same stage whose previous events only differ in $\{X_1, \ldots, X_{i-1}\} \backslash pa(X_i)$. For example, the statement $X_3 \perp\!\!\!\perp X_2 \mid X_1$ puts v_3 into a stage with v_4 and v_5 into a stage with v_6 (see Fig. 3). We hence have a one-to-one correspondence between the stages of the tree and the parent configurations $pa(x_i)$, $i = 2, 3, 4$, of the BN. Also, note that consequently a BN with no conditional independencies corresponds to a CEG where each node is in a separate stage. For our example we have the corresponding staged tree given in Fig. 3 with stages:

$$u_0 = \{v_0\},\ u_1 = \{v_1\},\ u_2 = \{v_2\},\ u_3 = \{v_3, v_4\},\ u_4 = \{v_5, v_6\},$$

$$u_5 = \{v_7, v_{10}, v_{13}, v_{16}\},\ u_6 = \{v_8, v_{11}, v_{14}, v_{17}\},\ u_7 = \{v_9, v_{12}, v_{15}, v_{18}\}.$$

We have that the conditional independence statement $X_3 \perp\!\!\!\perp X_2 \mid X_1$ is described by u_3 and u_4 and the conditional independence statement $X_4 \perp\!\!\!\perp X_1, X_2 \mid X_3$ by u_5, u_6 and u_7.

From the staged tree we can then simply construct the CEG. In our example the stages and positions coincide, and hence the vertices of the CEG are

$$w_0 = \{v_0\},\ w_1 = \{v_1\},\ w_2 = \{v_2\},\ w_3 = \{v_3, v_4\},\ w_4 = \{v_5, v_6\},\ w_5 = \{v_7, v_{10}, v_{13}, v_{16}\},$$

$$w_6 = \{v_8, v_{11}, v_{14}, v_{17}\},\ w_7 = \{v_9, v_{12}, v_{15}, v_{18}\},\ w_\infty = \{v_{19}, \ldots, v_{42}\},$$

with the corresponding graph given in Fig. 4.

In general, to move from a BN to a CEG we can go through each of the conditional independence statements $X_i \perp\!\!\!\perp \{X_1, \ldots, X_{i-1}\} \backslash pa(X_i) \mid pa(X_i)$ for $i = 2, 3, \ldots$ determining at each step the vertices describing X_i and colouring them accordingly to display the different stages.

It can be seen from Fig. 4 that the CEG of a BN is very symmetric. The CEG, however, contains a much richer set of conditional independence statements which can be read from the graph, although these often entail conditional independence statements about functions of subvectors of the original problem. For example, in circumstances like those described in the paper, we define a cut as a collection of positions where each unit passes through exactly one of the positions in the cut. Let Z denote a random variable denoting which of the positions of a cut a unit passes through. Then, given Z, what happens to the unit before arriving at the cut is independent of what happens after the unit leaves the cut. Thus, in our example, the cut $U = (w_3, w_4)$ is described by $Z = X_1$ and hence we can deduce by the above that $X_3 \perp\!\!\!\perp X_2 \mid X_1$. A more detailed description of this is given by Smith and Anderson [24]. Reading all types of entailed conditional independence statements associated with a CEG is straightforward but rather technical and the interested reader is referred to Ref. [26]. The increased generality of the CEG, however, allows us to search a much larger space of graphically explanatory models to find higher scoring models with less symmetry. We note that the loss of symmetry does not make the resulting fitted model any less interpretable, as will be illustrated in the next section.

5. Scoring Bayesian Networks and Chain Event Graphs

When searching the CEG space we can find the most probable BN derived previously, as the BN is a subclass of CEGs. However, if we consider the BN structures in Fig. 2 as well as the deductions we made by Fergusson et al. [9], it seems likely that the CEG will combine certain vertices into stages and positions in an asymmetric way resulting in a better model score. For example, it may be that a child from a lower social background and a low number of life events has the same probability of admissions as a child from a family with a higher social background and an average number of life events.

To score all possible CEGs we set up a scoring method for CEGs corresponding directly to the BD-metric for BNs developed in [12] to allow for a direct comparison between BNs and CEGs. Freeman and Smith [10] show that the axioms in [12] required for the BD-metric and the set up of the hyperparameters of the priors in the BN can be extended in a natural way to the CEG. They prove that under the assumptions that the stage priors are a priori independent and are identical for equivalent stages in different CEG structures, Dirichlet priors on the conditional probability vectors of the staged tree are inevitable. Further priors across models can then be deduced by summing the hyperparameters of the Dirichlet distribution when two stages are merged.

The BD-metric for CEGs is then the posterior probability of a CEG structure C given a complete multinomial sample D which is given explicitly as

$$P(C|D) = \frac{P(C)}{P(D)} P(D|C) = \frac{P(C)}{P(D)} \prod_{u \in J(T)} \frac{\Gamma(\alpha_u)}{\Gamma(\alpha_u + N_u)} \prod_{k=1}^{r_u} \frac{\Gamma(\alpha_{uk} + N_{uk})}{\Gamma(\alpha_{uk})}, \tag{1}$$

with $\alpha_u = \sum_k \alpha_{uk}$ and $N_u = \sum_k N_{uk}$, where α_{uk} are the parameters of the Dirichlet distribution describing the probability of going from a vertex in stage u to a vertex k and, similarly, $N_{uk}, k = 1, \ldots, r_u$, is the number of times we observe an individual going from a vertex in stage u to a vertex k. $\Gamma(.)$ is the usual Gamma function.

As for the BN, a uniform prior was given to the root-to-leaf paths of the finest partition of the CEG, in which all non-leaf vertices are in a separate stage. To enable a direct comparison the equivalent sample size was also chosen to be 3, as for the BNs. Assuming that different CEG structures are a priori equally likely, we can score and compare CEGs by looking simply at their marginal likelihoods $P(D|C)$. Similarly, we can hence look at the Bayes Factor scores (the ratio of the marginal likelihood) between BNs and CEGs to compare the model fit. We note that some care is needed when the priors are not chosen to be equally likely as the class of CEGs is larger than the class of BNs.

Since the CEG model space is far larger than the space of BN structures it is not feasible to perform an exhaustive search in all but the simplest case. In our example we therefore implemented a Bayesian Agglomerate Clustering (AHC) Algorithm developed for CEGs in [10]. The algorithm starts at the finest partition of the CEG, where each non-leaf vertex is in a separate stage. It then quickly searches over the model space by finding at every step the two stages, which, when merged, provide the highest CEG score, given by the BD-metric for CEGs. The new model's score and the corresponding CEG structure are recorded at every step. The algorithm stops once the coarsest partition of the CEG has been reached, where all non-leaf vertices with the same topology are in a single stage. The CEG with the highest overall score is then selected. We note that the algorithm favours simpler models and it can be shown that, when data is sparse, the algorithm tends to gather situations into the same stage.

The AHC algorithm found the highest scoring CEG structure to be the graph in Fig. 5. The corresponding stages are:

$$u_0 = \{v_0\},\ u_1 = \{v_1\},\ u_2 = \{v_2\},\ u_3 = \{v_3, v_4, v_5\},\ u_5 = \{v_6\},$$
$$u_6 = \{v_7, v_{10}\},\ u_8 = \{v_8, v_{11}, v_{13}, v_{14}, v_{16}\},\ u_9 = \{v_9, v_{12}, v_{15}, v_{17}, v_{18}\},$$

where u_3 is split into two positions w_3 and w_4 in Fig. 5.

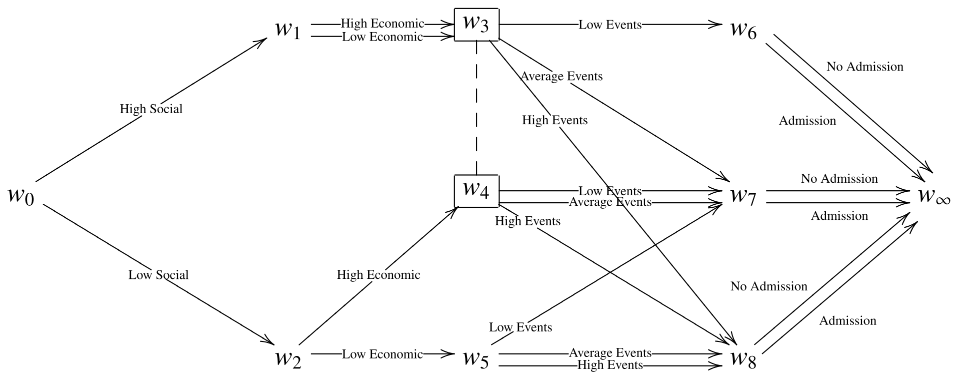

Fig. 5. CEG found by the AHC algorithm.

Table 3
Conditional probability table for CEG.

Stage	Conditional probability vector	
u_0	$(P(X_1 = \text{High}), P(X_1 = \text{Low}))$	$(0.569, 0.431)$
u_1	$P(X_2 = \text{High}\|u_1) + P(X_2 = \text{Low}\|u_1)$	1
u_2	$(P(X_2 = \text{High}\|u_2), P(X_2 = \text{Low}\|u_2))$	$(0.122, 0.878)$
u_3	$(P(X_3 = \text{Low}\|u_3), P(X_3 = \text{Average}\|u_3), P(X_3 = \text{High}\|u_3))$	$(0.461, 0.343, 0.196)$
u_5	$(P(X_3 = \text{Low}\|u_5), P(X_3 = \text{Average}\|u_5), P(X_3 = \text{High}\|u_5))$	$(0.220, 0.312, 0.468)$
u_6	$(P(X_4 = \text{No admission}\|u_6), P(X_4 = \text{Admission}\|u_6))$	$(0.910, 0.090)$
u_7	$(P(X_4 = \text{No admission}\|u_7), P(X_4 = \text{Admission}\|u_7))$	$(0.794, 0.206)$
u_8	$(P(X_4 = \text{No admission}\|u_8), P(X_4 = \text{Admission}\|u_8))$	$(0.743, 0.257)$

The graph illustrates several asymmetries which allow us to draw a number of conclusions on the combined effect of the three covariates on the hospital admissions that are not obtainable from the BN. Below we give a detailed account of the way we can interpret the topology of the CEG.

1. The social background appears to have an effect on the economic situation as suggested by the BN ($w_0 \to w_1$, $w_0 \to w_2$).
2. The economic situation seems to have no effect on the number of life events for families from a higher social background ($w_1 \to w_3$). However, in a family from a lower social background the economic situation seems to affect the number of life events that occur ($w_2 \to w_4, w_2 \to w_5$), as might be expected. We note that w_1 and its emanating edges could be removed in Fig. 5, with the edge emanating from w_0 leading directly into w_3.
3. Children from a family of high social background and a low number of life events, independent of the economic situation, are in a separate position and hence have a different probability of admissions to the other individuals ($w_3 \to w_6$).
4. Children from socially advantaged families with an average number of life events are in the same position as children from socially disadvantaged families with a high economic situation and a low or average number of life events, as are children from a low economic situation with a low number of life events ($w_3 \to w_7, w_4 \to w_7, w_5 \to w_7$).
5. All individuals with a high number of life events are in the same position irrespective of their social or economic background. Further, an individual from a low social and economic background with only an average number of life events is also in this position ($w_3 \to w_8, w_4 \to w_8, w_5 \to w_8$).

When comparing this CEG structure with the MAP BN structure, we obtain a Bayes Factor of 79,698 in favour of the CEG, providing very strong evidence that the additional flexibility of the CEG enables us to find a strongly preferable model. Table 3 gives the vectors of predictive probabilities associated with each stage.

Of particular interest are the three final positions w_6, w_7 and w_8 as these give an interpretation of the effect of a combination of variables on the hospital admissions as described in points 3 to 5 above. The CEG lets us trace the different paths the individuals can take ending up in one of these three positions. We have that the predicted probability of a hospital admission for individuals that reach position w_6 is 0.09. Further we predict an admission probability of 0.206 for individuals that reach position w_7 and 0.257 for the individuals at position w_8. We note that while a high number of life events forces the individuals into position w_8 with the highest admission probability, an individual from a low social background will never reach position w_6 even when he only has a low number of life events and his economic situation is high. The table of predictive probabilities further illustrates that a child from a low social background is more likely to also have a low economic background. Similarly, a child from a low social and economic background has a predictive probability of 0.461 to have a high number of life events and 0.196 for a low number of life events, while for the remaining children these probabilities are 0.220 and 0.468, respectively. It is also interesting to compare these probabilities with the probability vectors of the MAP BN in Table 2. Our results support the conclusions of Fergusson et al. [9], that the effect of life events on admissions is strongest. However, the CEG further explains explicitly the way in which the social background and the economic situation may have an additional effect on hospital admissions.

If we switched the life events and the economic situation in the ordering of the event tree we obtain a MAP CEG structure with the same final three positions, w_5, w_6 and w_7. More explicitly, only the way in which the events may affect the economic situation is novel, while the overall conclusions on hospital admissions remain the same. In this situation we would have that children from a high social background with a low or average number of life events have the same distribution for the economic situation. Also, children from a high social background with a high number of life events or from a low social background with a low number of life events are in one stage and, similarly, families from a low social background with an average or high number of life events are in the same stage. This CEG only scores very slightly less than the CEG discussed previously, with a Bayes Factor of 1.27 in favour of the previous CEG but with a Bayes Factor of 70,686 in comparison to the MAP BN.

We further note that the CEG provides a useful improvement to Generalised Linear Models. When carrying out a logistic regression on the effect of the social background, economic situation and life events on the probability of hospital admission we would need to include all possible two-way and three-way interaction terms to be able to make inference on the combined effect of the covariates on the outcome. Given the parameter estimates of the regression model we could

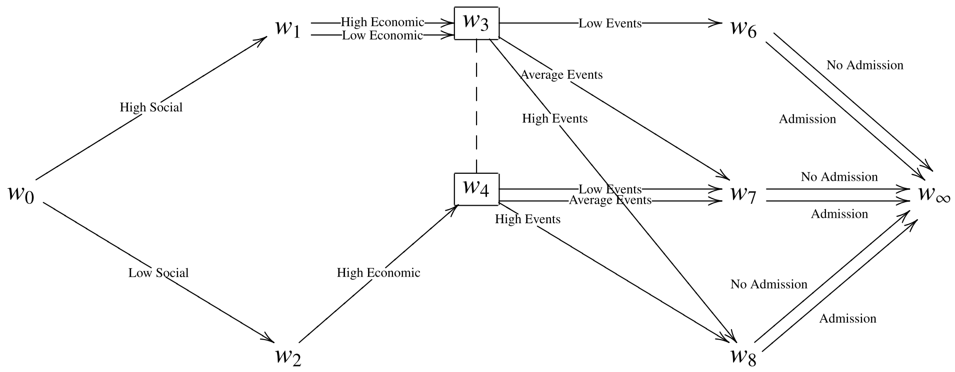

Fig. 6. Manipulated CEG.

then calculate and compare the estimated effect of each combination of covariates given the baseline. However, this easily becomes infeasible as the number of covariates increases. To further determine the effect of the social background on the economic situation and the effect of social background and economic situation on life events we would need to look at two further logistic regressions. The CEG combines all this information within the topology of its graph and shows directly which combination of covariates have similar effects on the admission probability. In addition to this the AHC algorithm automatically determines when the effects of two different combinations of covariates can be interpreted as being the same (the vertices are in the same stage) taking into account the complexity of the model and the number of counts in each category.

6. The causal interpretation of Chain Event Graphs

We further illustrate that just as with BNs these graphs can be linked to causal hypotheses about the likely effect of interventions which can then be tested in the future. In particular, the CEG lets us make refined causal statements about the hospital admissions by allowing for possible interventions to be asymmetric in the sense that we can intervene at a particular position rather than only on a variable. The simplest form of intervention forces an individual that reaches a position w along a particular edge, $w \to w^*$, say, and hence the conditional probability vector at w has one entry equal to 1 and zeros everywhere else. Thwaites et al. [28] then define, analogous to Pearl [20], a Causal CEG to be a CEG which, under the manipulation at a position w, obeys the intervention formula that $P(w \to w^*|w) = 1$ and hence $P(w \to w'|w) = 0$ for $w' \neq w^*$. All other probabilities are as in the unmanipulated tree. The manipulated CEG can be drawn by deleting all paths emanating from w that do not go along the required edge and we can calculate the effect of the intervention directly from the table of conditional probability vectors.

In our example, if we are prepared to read the graph causally, the effect of giving only families from a low social background financial aid corresponds to forcing all individuals that reach w_2 along the edge $w_2 \to w_4$. The manipulated CEG is given in Fig. 6.

The predicted probability of hospital admissions for families from a low social background and an average number of life events given this intervention is improved to 0.206 as can be found in Table 3. Further we can read from the table that the intervention reduces the predicted probability of a high number of life events in socially disadvantaged families from 0.468 to 0.196, giving an improvement in the probability of admissions mediated through the life events. We further emphasise here the point made by Shafer [22] that causal assumptions should be inferred from tree-like structures as these naturally respect a causal ordering and in this sense we can see directly from the graph of the CEG and the associated conditional probability vectors at which position we may want to intervene.

7. Discussion

In this paper we have demonstrated through data on childhood hospital admissions that, whilst a BN search can be very useful for finding good explanations and providing a graphical framework for feeding back the analysis, the CEG provides useful refinements to an initial BN study. This is not only apparent in the significantly high Bayes Factor of the derived CEG and the MAP BN structure but also in its expressiveness for the client.

Although the CEGs structural syntax is closely linked to PDGs it is a more general class of models due to the additional colouring of the CEG when two positions are in the same stage. Jaeger [14] showed that PDGs and BNs are incomparable regarding the conditional independence statements they encode and hence that the BN is not a subclass of the PDG. In contrast to this we have demonstrated in this paper that any BN can be written as a CEG and therefore, unlike for the PDG, the techniques designed for BNs could be easily extended to CEGs to enhance an initial BN search with a subsequent CEG analysis. In particular, the MAP model given in Fig. 5 could not be represented as a PDG, due to the colouring of the CEG with respect to w_4 and w_5, which gives additional information on the effect of the social background and economic situation on the life events.

Please cite this article in press as: L.M. Barclay et al., Refining a Bayesian Network using a Chain Event Graph, International Journal of Approximate Reasoning (2013), http://dx.doi.org/10.1016/j.ijar.2013.05.006

There are several features of interest which could not be addressed here. In particular, the dynamic structure of the data could be used to provide further insight. Current research is being carried out on the development of a dynamic version of the CEG. The flexibility of the CEG of allowing the nodes of the probability tree to be combined in an asymmetric way can be exploited even more in the dynamic setting as nodes from different time points can be merged. With respect to health studies, such as the Christchurch Health and Development Study, these models would be very useful for describing the development of changing health status of the children due to family factors occurring at different points in time.

A further point of interest to be addressed in the future is the development of improved model selection algorithms to enable more efficient selection procedures. We are currently developing a dynamic programming algorithm for CEGs, based on [23], which is able to search the CEG space for problems with up to ten variables. Parallel to this the use of CEGs for more complex problems is being considered. The development of an ordinal CEG allows for an improved graphical representation by listing the positions in descending order according to a variable of interest. In addition to this it is often possible to reduce the original CEG structure by defining new variables resulting from the dependence structure of the variables depicted in the topology of the CEG. These aspects are discussed in detail in a subsequent paper [2].

However, even when the analyses are further refined in the ways described above, the point illustrated in this paper still appears to hold: The CEG exhibits significant advantages in the application of graphical models to population studies like those of our example. In particular, in a large BN study, the dependence structure between a subset of variables may not be clear and further refinements may be desirable. This can easily be obtained through a CEG analysis on a selected subset of the variables. We would therefore encourage BN modellers to explore the potential of enhancing a BN search with a subsequent CEG analysis either on the full model or simply on a subset of the variables of the BN.

Acknowledgements

The authors would like to thank John Horwood and the CHDS research group for providing the data set. The authors would also like to thank the reviewers whose comments greatly enhanced the paper. One of the authors was funded by the Engineering and Physical Sciences Research Council [grant number EP/P50578X/1].

References

[1] Christchurch health and development study, http://www.otago.ac.nz/christchurch/research/healthdevelopment/, Accessed: February 2012.
[2] L. Barclay, J. Hutton, J. Smith, Chain event graphs for informed missingness, CRiSM Research Report 12-17, 2012.
[3] A. Beautrais, D. Fergusson, F. Shannon, Life events and childhood morbidity: A prospective study, Pediatrics 70 (1982) 935–940.
[4] S. Bøttcher, C. Dethlefsen, deal: A package for learning Bayesian networks, Journal of Statistical Software 8 (2003) 1–40.
[5] C. Boutilier, N. Friedman, M. Goldszmidt, D. Koller, Context-specific independence in Bayesian networks, in: Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence Portland, Oregon, 1996, pp. 115–123.
[6] A. Cano, M. Gómez-Olmedo, S. Moral, C.B. Pérez-Ariza, A. Salmerón, Learning recursive probability trees from probabilistic potentials, International Journal of Approximate Reasoning 53 (9) (2012) 1367–1387.
[7] R. Cowell, A. Dawid, S. Lauritzen, D. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer Verlag, 2007.
[8] D. Fergusson, M. Dimond, L. Horwood, F. Shannon, The utilisation of preschool health and education services, Social Science & Medicine 19 (1984) 1173–1180.
[9] D. Fergusson, L. Horwood, F. Shannon, Social and family factors in childhood hospital admission, Journal of Epidemiology and Community Health 40 (1986) 50.
[10] G. Freeman, J. Smith, Bayesian map model selection of chain event graphs, Journal of Multivariate Analysis 102 (2011) 1152–1165.
[11] N. Friedman, M. Goldszmidt, Learning Bayesian networks with local structure, NATO ASI Series D Behavioural and Social Sciences 89 (1998) 421–460.
[12] D. Heckerman, D. Geiger, D. Chickering, Learning Bayesian networks: The combination of knowledge and statistical data, Machine Learning 20 (1995) 197–243.
[13] T. Holmes, R. Rahe, The social readjustment rating scale, Journal of Psychosomatic Research 11 (1967) 213–218.
[14] M. Jaeger, Probabilistic decision graphs – combining verification and AI techniques for probabilistic inference, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (2004) 19–42.
[15] M. Jaeger, J.D. Nielsen, T. Silander, Learning probabilistic decision graphs, International Journal of Approximate Reasoning 42 (2006) 84–100.
[16] K. Korb, A. Nicholson, Bayesian Artificial Intelligence, vol. 1, CRC Press, 2004.
[17] R. Neapolitan, Learning Bayesian Networks, Pearson–Prentice Hall, Upper Saddle River, NJ, 2004.
[18] J.D. Nielsen, R. Rumí, A. Salmerón, Structural-EM for learning PDG models from incomplete data, International Journal of Approximate Reasoning 51 (2010) 515–530.
[19] J.J. Oliver, Decision graphs – an extension of decision trees, in: Proceedings of the Fourth International Workshop on Artificial Intelligence and Statistics, 1993, pp. 343–350.
[20] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd edition, Cambridge University Press, Cambridge, 2009.
[21] D. Poole, N. Zhang, Exploiting contextual independence in probabilistic inference, Journal of Artificial Intelligence Research 18 (2003) 263–313.
[22] G. Shafer, The Art of Causal Conjecture, The MIT Press, 1996.
[23] T. Silander, P. Myllymäki, A simple approach for finding the globally optimal Bayesian network structure, in: Proceedings of the 22nd Conference on Artificial intelligence (UAI 2006), AUAI Press, 2006, pp. 445–452.
[24] J. Smith, P. Anderson, Conditional independence and chain event graphs, Artificial Intelligence 172 (2008) 42–68.
[25] P. Thwaites, Causal identifiability via chain event graphs, Artificial Intelligence 195 (2013) 291–315.
[26] P. Thwaites, J. Smith, Separation theorems for chain event graphs, CRiSM Research Report 11-09, 2011.
[27] P. Thwaites, J. Smith, R. Cowell, Propagation using chain event graphs, in: Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-08), AUAI Press, Corvallis, Oregon, 2008, pp. 546–553.
[28] P. Thwaites, J. Smith, E. Riccomagno, Causal analysis with chain event graphs, Artificial Intelligence 174 (2010) 889–909.
[29] S. Weich, J. Patterson, R. Shaw, S. Stewart-Brown, Family relationships in childhood and common psychiatric disorders in later life: Systematic review of prospective studies, The British Journal of Psychiatry 194 (2009) 392–398.